Contents list available at ScienceDrect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Hyper least squares fitting of circles and ellipses

Kenichi Kanatani[a,*], Prasanna Rangarajan[b]

[a] *Department of Computer Science, Okayama University, Okayama 700-8530, Japan*
[b] *Department of Electrical Engineering, Southern Methodist University, Dallas, TX 75205 U.S.A.*

ARTICLE INFO

Article history:
Received 10 May 2010
Received in revised form 28 December 2010
Accepted 29 December 2010
Available online 11 Janurary 2011

Keywords:
Least squares
Circle fitting
Ellipse fitting
Algebraic distance minimization
Error analysis
Bias removal

ABSTRACT

This work extends the circle fitting method of Rangarajan and Kanatani (2009) to accommodate ellipse fitting. Our method, which we call HyperLS, relies on algebraic distance minimization with a carefully chosen scale normalization. The normalization is derived using a rigorous error analysis of least squares (LS) estimators so that statistical bias is eliminated up to second order noise terms. Numerical evidence suggests that the proposed HyperLS estimator is far superior to the standard LS and is slightly better than the Taubin estimator. Although suboptimal in comparison to maximum likelihood (ML), our HyperLS does not require iterations. Hence, it does not suffer from convergence issues due to poor initialization, which is inherent in ML estimators. In this sense, the proposed HyperLS is a perfect candidate for initializing the ML iterations.

1. Introduction

Detecting circles and ellipses in images and computing their mathematical representation are the first step of many computer vision applications including industrial robotic operations and autonomous navigation (Hartley and Zisserman, 2004; Kanatani, 1993, 1996). Among many circle/ellipse fitting algorithms presented in the past, those regarded as the most accurate are methods based on maximum likelihood (ML), and various computational schemes have been proposed including the *FNS* (*Fundamental Numerical Scheme*) of Chojnacki et al. (2000), the *HEIV* (*Heteroscedastic Errors-in-Variable*) of Leedan and Meer (2000) and Matei and Meer (2006), and the *projective Gauss-Newton iterations* of Kanatani and Sugaya (2007). Strictly speaking, these schemes do not directly maximize the likelihood but minimize what is known as the *Sampson error* (Hartley and Zisserman, 2004), but it can be proved that the ML solution is obtained by running a few iterations of a Sampson error minimization algorithm (Kanatani and Sugaya, 2010). It has also been observed that the solution that minimizes the Sampson error agrees with the ML solution up to several significant digits (Kanatani and Sugaya, 2008). Efforts have also been made to add a posterior correction (Kanatani, 2006), but all ML-based methods already achieve the theoretical accuracy limit, called the *KCR lower bound* (Chernov and Lesort, 2004; Kanatani, 1996, 2008), up to high order noise terms. Hence, there is practically no room for further accuracy improvement. However, all ML-based methods have one drawback: iterations are required for nonlinear optimization, and the computation often fails to converge in the presence of large noise, depending on the accuracy of the initialization. Therefore, accurate algebraic methods that do not require iterations are very much desired, even though the solution may not be strictly optimal.

An "algebraic" method refers to minimizing an easy-to-minimize polynomial, called "algebraic distance". However, all algebraic methods have an inherent weakness: we need to impose a normalization to remove scale indeterminacy, and

* Corresponding author. Tel.: +81 86 251 8173; fax +81 86 251 8173.
E-mail addresses: kanatani@suri.cs.okayama-u.ac.jp (K. kanatani), prangara@mail.smu.edu (P. Rangarajan).

0167-9473/$–see front matter ©2011 Elsevier B.V. All rights reserved.
doi:10.10.16/j.csda.2010.12.012

the solution depends on the choice of the normalization. Al-Sharadqah and Chernov (2009) and Rangarajan and Kanatani (2009) exploited this freedom for fitting circles (see Al-Sharadqah and Chernov (2009) for comprehensive references of existing circle fitting algorithms) and optimized the normalization so that the solution has high accuracy. The aim of this paper is to generalize their approach to include ellipses as well.

In Secion 2, we describe the algebraic approach for circle/ellipse fitting and point out the importance of choosing a correct normalization. Section 3 presents a rigorous error analysis of algebraic solutions using the perturbation theory of Kanatani (2008). In Section 4, we derive concrete expressions for the covariance and bias of algebraic solutions in general and their specific forms for the standard least squares (LS) and the Taubin method, an algebraic method known to be highly accurate. Section 5 presents our method, which we call *hyper least squares*, or *HyperLS*, by deriving a normalization such that the solution is unbiased up to the second order noise terms. In Section 6, we provide numerical evidence which suggests that HyperLS performs far better than the standard LS and, although the difference is small, is even superior to the Taubin method. Thus, it is a best candidate for the initialization of ML iterations. In Section 7, we conclude.

2. Algebraic fitting

A circle is represented by

$$A(x^2+y^2)+2f_0(Dx+Ey)+f_0^2F=0, \tag{1}$$

and an ellipse by

$$Ax^2+2Bxy+Cy^2+2f_0(Dx+Ey)+f_0^2F=0, \tag{2}$$

where f_0 is a scale constant that has the order of the x and y coordinates of the data; without it, finite precision numerical computation would incur serious accuracy loss when x and y are in the order of 100–1000, a typical situation in image processing applications. Letting f_0 be of the order of x and y is equivalent to scaling x and y into the range ± 1. For interested readers, a brief discussion is given in Appendix A on how the solution would change by altering the value of f_0 using a hypothetical computer with infinite precision.

Our task is to compute the coefficients A, ..., F so that the circle of (1) or the ellipse of (2) fits the given points (x_α, y_α), $\alpha = 1$, ..., N, as closely as possible. The least squares (LS) estimator minimizes the *algebraic distance*

$$J=\frac{1}{N}\sum_{\alpha=1}^{N}\Big(A(x_\alpha^2+y_\alpha^2)+2f_0(Dx_\alpha+Ey_\alpha)+f_0^2F\Big)^2 \tag{3}$$

for circles and

$$J=\frac{1}{N}\sum_{\alpha=1}^{N}\Big(Ax_\alpha^2+2Bx_\alpha y_\alpha+Cy_\alpha^2+2f_0(Dx_\alpha+Ey_\alpha)+f_0^2F\Big)^2 \tag{4}$$

for ellipses. This approach is also known as the *direct linear transformation* (*DLT*) (Hartley and Zisserman, 2004). Evidently, (3) and (4) are minimized by $A=\cdots=F=0$ if no restriction is placed on A, ..., F. In an effort to avoid the trivial solution, many forms of normalization have been proposed for ellipse fitting including

$$F=1, \tag{5}$$

$$A+C=1, \tag{6}$$

$$A^2+B^2+C^2+D^2+E^2+F^2=1, \tag{7}$$

$$A^2+B^2+C^2+D^2+E^2=1, \tag{8}$$

$$A^2+2B^2+C^2=1, \tag{9}$$

$$AC-B^2=1. \tag{10}$$

The normalization (5) reduces minimization of (4) to simultaneous linear equations (Albano, 1974; Cooper and Yalabik, 1976; Rosin, 1993). However, the ellipse (2) with (5) cannot pass through the origin $(0,0)$. The use of (6) remedies this (Gander et al., 1995; Porrill, 1990; Rosin and West, 1995). The most frequently used is (7) (Paton, 1970) [1], but some authors use (8) (Gnanadesikan, 1977). The use of (9) imposes invariance to coordinate transformations in the sense that the ellipse fitted after the coordinate system is translated and rotated is the same as the originally fitted ellipse translated and rotated afterwards (Bookstein, 1979). In this respect, (6) and (10) also have that invariance. The use of (10) prevents (2) from representing a parabola ($AC-B^2=0$) or a hyperbola ($AC-B^2<0$) (Fitzgibbon et al., 1999). In this paper, we do not exclude nonellipse solutions but derive a method that computes A, ..., F as closely to their true values as

[1] Some authors write an ellipse as $Ax^2+Bxy+Cy^2+Dx+Ey+F=0$. The meaning of (7) changes accordingly, but we ignore such differences; no significant consequence would result.

possible. For circles, too, Al-Sharadqah and Chernov (2009) pointed out that many existing methods, including those of Kåsa (1976), Pratt (1987), and Taubin (1991), each derived rather heuristically, can be characterized by the difference in scale normalization. The purpose of this paper is to find theoretically the "best" normalization.

If we define $\boldsymbol{\xi}$ and $\boldsymbol{\theta}$ by

$$\boldsymbol{\xi} = (x^2 + y^2, 2f_0x, 2f_0y, f_0^2)^\top, \qquad \boldsymbol{\theta} = (A, D, E, F)^\top \tag{11}$$

for circles and

$$\boldsymbol{\xi} = (x^2, 2xy, y^2, 2f_0x, 2f_0y, f_0^2)^\top, \qquad \boldsymbol{\theta} = (A, B, C, D, E, F)^\top \tag{12}$$

for ellipses, (1) and (2) can be expressed in the compact form

$$(\boldsymbol{\xi}, \boldsymbol{\theta}) = 0, \tag{13}$$

where and hereafter we denote the inner product of vectors $\mathbf{a}$ an $\mathbf{b}$ by $(\mathbf{a}, \mathbf{b})$. Let $\boldsymbol{\xi}_\alpha$ be the value of $\boldsymbol{\xi}$ for (x_α, y_α). Then, (3) and (4) can be expressed as

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{\theta}, \boldsymbol{\xi}_\alpha)^2 = \frac{1}{N}\sum_{\alpha=1}^{N}\boldsymbol{\theta}^\top\boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top\boldsymbol{\theta} = (\boldsymbol{\theta}, \mathbf{M}\boldsymbol{\theta}), \tag{14}$$

where we define the matrix $\mathbf{M}$ as follows:

$$\mathbf{M} = \frac{1}{N}\sum_{\alpha=1}^{N}\boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top. \tag{15}$$

As mentioned before, (14) is trivially minimized by $\boldsymbol{\theta} = \mathbf{0}$ unless some scale normalization is imposed on $\boldsymbol{\theta}$. Following Al-Sharadqah and Chernov (2009) and Rangarajan and Kanatani (2009), we consider the following class of normalizations

$$(\boldsymbol{\theta}, \mathbf{N}\boldsymbol{\theta}) = c \tag{16}$$

with some symmetric matrix $\mathbf{N}$ for a nonzero constant c. In (16), $\boldsymbol{\theta}$ is the optimization parameter and $\mathbf{N}$ is an unknown matrix to be determined, while c is a constant fixed for the problem. We need not specify the value of c, because $\mathbf{N}$ is unknown. In fact, (16) can be written as $(\boldsymbol{\theta}, (\mathbf{N}/c)\boldsymbol{\theta}) = 1$, and we may alternatively determine the unknown $\mathbf{N}' = \mathbf{N}/c$ instead of $\mathbf{N}$. However, the form of (16) with the value and its signature of c unspecified is more convenient in our analysis.

Note that (7), (8), (9), and (10) are in the form of (16) as are, but (5) and (6) can also be included as $F^2 = 1$ and $(A+C)^2 = 1$. For (5)–(9), the matrix $\mathbf{N}$ is positive semidefinite (positive definite for (7)) but not for (10). In the following, we allow $\mathbf{N}$ to be nondefinite (i.e., neither positive nor negative definite), so the constant c in (16) is not necessarily positive. Given the matrix $\mathbf{N}$, the standard treatment of algebraic fitting goes as follows. As is well known, the solution $\boldsymbol{\theta}$ that minimizes (14) subject to (16), if it exists, is obtained as the solution of the generalized eigenvalue problem

$$\mathbf{M}\boldsymbol{\theta} = \lambda\mathbf{N}\boldsymbol{\theta}. \tag{17}$$

If there is no noise in the data, we have $(\boldsymbol{\theta}, \boldsymbol{\xi}_\alpha) = 0$ for all α. Hence, (15) implies $\mathbf{M}\boldsymbol{\theta} = \mathbf{0}$, so $\lambda = 0$. If $\mathbf{N}$ is positive definite or semidefinite, the eigenvalue λ is positive in the presence of noise. The corresponding solution is obtained as the eigenvector $\boldsymbol{\theta}$ for the smallest λ. Let us call the popular method of letting $\mathbf{N} = \mathbf{I}$ (the use of (7) for ellipses) the *standard LS*. In this case, (17) becomes an ordinary eigenvalue problem

$$\mathbf{M}\boldsymbol{\theta} = \lambda\boldsymbol{\theta}, \tag{18}$$

and the solution is the unit eigenvector $\boldsymbol{\theta}$ of $\mathbf{M}$ for the smallest eigenvalue λ.

This is the traditional treatment of algebraic fitting, but the situation is slightly different in our case. Here, $\mathbf{N}$ is not yet given and can be nondefinite; the eigenvalues may not be all positive. So, we face the problem of which of the eigenvalues and eigenvectors of (17) to choose as a solution. In the following, we do perturbation analysis of (17) by assuming that $\lambda \approx 0$ (Kanatani, 2008) and choose the solution to be the eigenvector $\boldsymbol{\theta}$ for the λ with the smallest absolute value. We also regard (17) as the definition of our "algebraic method", rather than (14) and (16). This is because, while (17) always has a solution, (14) may not be minimized subject to (16) by a finite $\boldsymbol{\theta}$. This can occur, for example, when the contour of $(\boldsymbol{\theta}, \mathbf{M}\boldsymbol{\theta})$, which is a hyperellipsoid in the 9-D space of $\boldsymbol{\theta}$, happens to be elongated in a direction in the null space of $\mathbf{N}$. Then, the minimum of $(\boldsymbol{\theta}, \mathbf{M}\boldsymbol{\theta})$ could be reached in the limit of $\|\boldsymbol{\theta}\| \to \infty$. Theoretically, such an anomaly can always occur because $\mathbf{M}$ is a random variable defined by noisy data, and if the probability of such an occurrence is nearly 0, it may still lead to $E[\|\hat{\boldsymbol{\theta}}\|] = \infty$ (Cheng and Kukush, 2006). Once the problem is converted to (17), for which eigenvectors $\boldsymbol{\theta}$ have scale indeterminacy, we can adopt normalization $\|\boldsymbol{\theta}\| = 1$ rather than (16). Then, the solution $\boldsymbol{\theta}$ is always a unit vector.

3. Error analysis

Before proceeding to the error analysis of algebraic fitting, we need to introduce a statistical model of observation. We regard each (x_α, y_α) as perturbed from its true position $(\bar{x}_\alpha, \bar{y}_\alpha)$ by $(\Delta x_\alpha, \Delta y_\alpha)$, where Δx_α and Δy_α are Gaussian

random variables of mean 0 and standard deviation σ. We do not impose any restrictions on the true positions $(\bar{x}_\alpha, \bar{y}_\alpha)$ except that they should lie on a (true) circle/ellipse. This is known as a *functional* model. We could also introduce some statistical model according to which the true positions $(\bar{x}_\alpha, \bar{y}_\alpha)$ are sampled. Then, the model is called *structural*. This distinction is crucial when we consider limiting processes in the following sense (see Kanatani (2008) for more detailed discussions about this). The traditional statistical analysis mainly focuses on the asymptotic behavior as the number of observations increases to ∞. This is based on the reasoning that the mechanism underlying noisy observations would better reveal itself as the number of observations increases (the law of large numbers) while the number of available data is limited in practice. So, the estimation accuracy vs. the number of data is a major concern, and for ellipse fitting, too, a *consistent* estimator in the sense that it produces the true solution in the limit $N \to \infty$ of the number N of points has been sought (Kukush et al., 2004). In image processing applications, however, we observe a single set of image data, from which we do inference. The accuracy of the inference deteriorates as the noise in the data increases, so the inference accuracy vs. the noise level is a major concern. Usually, the noise is very small, often subpixel levels. In view of this, it has been pointed out that the "consistency" of estimators should more appropriately be defined by the behavior in the limit $\sigma \to 0$ of the noise level σ (Chernov and Lesort, 2004; Kanatani, 2008).

In this paper, we are interested in image processing applications, focusing on the perturbation analysis in the noise level $\sigma \approx 0$ for a fixed number N of points on the circle/ellipse. Thus, the functional model suits our purpose. If we were to analyze the error behavior in the limit of $N \to \infty$, a model that specifies how the data positions increase on the circle/ellipse, e.g., whether the distribution is uniform or not, would be necessary, and the derivation of consistent estimators for $N \to \infty$ (e.g., Kukush et al. Kukush et al. (2004)) is based on such an assumption.

Under our setting, the observation vector $\boldsymbol{\xi}_\alpha$ can be expressed as the sum

$$\boldsymbol{\xi}_\alpha = \bar{\boldsymbol{\xi}}_\alpha + \Delta_1\boldsymbol{\xi}_\alpha + \Delta_2\boldsymbol{\xi}_\alpha, \tag{19}$$

where $\bar{\boldsymbol{\xi}}_\alpha$ is the true value of $\boldsymbol{\xi}_\alpha$, and $\Delta_1\boldsymbol{\xi}_\alpha$, and $\Delta_2\boldsymbol{\xi}_\alpha$ are the noise terms of the first and the second order, respectively. For circles, we see from (11) that

$$\Delta_1\boldsymbol{\xi}_\alpha = (2\bar{x}_\alpha\Delta x_\alpha + 2\bar{y}_\alpha\Delta y_\alpha, 2f_0\Delta x_\alpha, 2f_0\Delta y_\alpha, 0)^\top, \qquad \Delta_2\boldsymbol{\xi}_\alpha = (\Delta x_\alpha^2 + \Delta y_\alpha^2, 0, 0, 0)^\top. \tag{20}$$

For ellipses, we see from (12) that

$$\begin{aligned}\Delta_1\boldsymbol{\xi}_\alpha &= (2\bar{x}_\alpha\Delta x_\alpha, 2\bar{x}_\alpha\Delta y_\alpha + 2\bar{y}_\alpha\Delta x_\alpha, 2\bar{y}_\alpha\Delta y_\alpha, 2f_0\Delta x_\alpha, 2f_0\Delta y_\alpha, 0)^\top,\\ \Delta_2\boldsymbol{\xi}_\alpha &= (\Delta x_\alpha^2, 2\Delta x_\alpha\Delta y_\alpha, \Delta y_\alpha^2, 0, 0, 0)^\top.\end{aligned} \tag{21}$$

We define the covariance matrix of $\boldsymbol{\xi}_\alpha$ by $V[\boldsymbol{\xi}_\alpha] = E[\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top]$, where $E[\,\cdot\,]$ denotes expectation. We regard Δx_α and Δy_α as independent Gaussian variables of mean 0 and standard deviation σ. Hence, $V[\boldsymbol{\xi}_\alpha] = \sigma^2 V_0[\boldsymbol{\xi}_\alpha]$, where

$$V_0[\boldsymbol{\xi}_\alpha] = \begin{pmatrix} 4(\bar{x}_\alpha^2 + \bar{y}_\alpha^2) & 2f_0\bar{x}_\alpha & 2f_0\bar{y}_\alpha & 0 \\ 2f_0\bar{x}_\alpha & f_0^2 & 0 & 0 \\ 2f_0\bar{y}_\alpha & 0 & f_0^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad V_0[\boldsymbol{\xi}_\alpha] = 4\begin{pmatrix} \bar{x}_\alpha^2 & \bar{x}_\alpha\bar{y}_\alpha & 0 & f_0\bar{x}_\alpha & 0 & 0 \\ \bar{x}_\alpha\bar{y}_\alpha & \bar{x}_\alpha^2 + \bar{y}_\alpha^2 & \bar{x}_\alpha\bar{y}_\alpha & f_0\bar{y}_\alpha & f_0\bar{x}_\alpha & 0 \\ 0 & \bar{x}_\alpha\bar{y}_\alpha & \bar{y}_\alpha^2 & 0 & f_0\bar{y}_\alpha & 0 \\ f_0\bar{x}_\alpha & f_0\bar{y}_\alpha & 0 & f_0^2 & 0 & 0 \\ 0 & f_0\bar{x}_\alpha & f_0\bar{y}_\alpha & 0 & f_0^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{22}$$

for circles and ellipses, respectively. Here, we have noted that $E[\Delta x_\alpha] = E[\Delta y_\alpha] = 0$, $E[\Delta x_\alpha^2] = E[\Delta y_\alpha^2] = \sigma^2$, and $E[\Delta x_\alpha\Delta y_\alpha] = 0$ according to our assumption. The method of Taubin (1991) is to use as $\mathbf{N}$ the matrix

$$\mathbf{N}_{\mathrm{T}} = \frac{1}{N}\sum_{\alpha=1}^{N} V_0[\boldsymbol{\xi}_\alpha]. \tag{23}$$

This matrix contains the true values $(\bar{x}_\alpha, \bar{y}_\alpha)$, which are replaced by the observations (x_α, y_α) in actual computation.

Substituting (19) into (15), we have

$$\mathbf{M} = \frac{1}{N}\sum_{\alpha=1}^{N}(\bar{\boldsymbol{\xi}}_\alpha + \Delta_1\boldsymbol{\xi}_\alpha + \Delta_2\boldsymbol{\xi}_\alpha)(\bar{\boldsymbol{\xi}}_\alpha + \Delta_1\boldsymbol{\xi}_\alpha + \Delta_2\boldsymbol{\xi}_\alpha)^\top = \bar{\mathbf{M}} + \Delta_1\mathbf{M} + \Delta_2\mathbf{M} + \cdots, \tag{24}$$

where $\cdots$ denotes noise terms of order three and higher. The matrix $\bar{\mathbf{M}}$ is the noise-free value of $\mathbf{M}$, and $\Delta_1\mathbf{M}$, and $\Delta_2\mathbf{M}$ are

$$\Delta_1\mathbf{M} = \frac{1}{N}\sum_{\alpha=1}^{N}\Big(\bar{\boldsymbol{\xi}}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top + \Delta_1\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\Big), \qquad \Delta_2\mathbf{M} = \frac{1}{N}\sum_{\alpha=1}^{N}\Big(\bar{\boldsymbol{\xi}}_\alpha\Delta_2\boldsymbol{\xi}_\alpha^\top + \Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top + \Delta_2\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\Big). \tag{25}$$

We expand the solution $\boldsymbol{\theta}$ and λ of (17) in the form

$$\boldsymbol{\theta} = \bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots, \qquad \lambda = \bar{\lambda} + \Delta_1\lambda + \Delta_2\lambda + \cdots, \tag{26}$$

where the barred terms are the noise-free values, and symbols Δ_1 and Δ_2 indicate the first and the second order noise terms, respectively. Substituting (24) and (26) into (17), we obtain

$$(\bar{\mathbf{M}} + \Delta_1\mathbf{M} + \Delta_2\mathbf{M} + \cdots)(\bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots) = (\bar{\lambda} + \Delta_1\lambda + \Delta_2\lambda + \cdots)\mathbf{N}(\bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots). \tag{27}$$

Note that $\mathbf{N}$ is a variable to be determined, not a given function of observations, so it is not expanded. Since we consider perturbations near the true values, the matrix $\mathbf{N}$ may turn out to involve them as in (23). In that event, we replace the true data values by their observations, as in the above Taubin case, and do an a posteriori analysis to see how this affects the accuracy. For the moment, we regard $\mathbf{N}$ as an unknown variable. Equating terms of the same order on both sides of (27), we obtain

$$\bar{\mathbf{M}}\bar{\boldsymbol{\theta}} = \bar{\lambda}\mathbf{N}\bar{\boldsymbol{\theta}}, \tag{28}$$

$$\bar{\mathbf{M}}\Delta_1\boldsymbol{\theta} + \Delta_1\mathbf{M}\bar{\boldsymbol{\theta}} = \bar{\lambda}\mathbf{N}\Delta_1\boldsymbol{\theta} + \Delta_1\lambda\mathbf{N}\bar{\boldsymbol{\theta}}, \tag{29}$$

$$\bar{\mathbf{M}}\Delta_2\boldsymbol{\theta} + \Delta_1\mathbf{M}\Delta_1\boldsymbol{\theta} + \Delta_2\mathbf{M}\bar{\boldsymbol{\theta}} = \bar{\lambda}\mathbf{N}\Delta_2\boldsymbol{\theta} + \Delta_1\lambda\mathbf{N}\Delta_1\boldsymbol{\theta} + \Delta_2\lambda\mathbf{N}\bar{\boldsymbol{\theta}}. \tag{30}$$

The noise-free values $\bar{\boldsymbol{\xi}}_\alpha$ and $\bar{\boldsymbol{\theta}}$ satisfy $(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{\theta}}) = 0$, so $\bar{\mathbf{M}}\bar{\boldsymbol{\theta}} = \mathbf{0}$. Hence, (28) implies $\bar{\lambda} = 0$. From the first equation in (25), we have $(\bar{\boldsymbol{\theta}}, \Delta_1\mathbf{M}\bar{\boldsymbol{\theta}}) = 0$. Computing the inner product of (29) with $\bar{\boldsymbol{\theta}}$ on both sides, we find that $\Delta_1\lambda = 0$. Multiplying (29) by the pseudoinverse $\bar{\mathbf{M}}^-$ from left, we have

$$\Delta_1\boldsymbol{\theta} = -\bar{\mathbf{M}}^-\Delta_1\mathbf{M}\bar{\boldsymbol{\theta}}, \tag{31}$$

where we have noted that because $\bar{\boldsymbol{\theta}}$ is the null vector of $\bar{\mathbf{M}}$ (i.e., $\bar{\mathbf{M}}\bar{\boldsymbol{\theta}} = \mathbf{0}$), the matrix $\bar{\mathbf{M}}^-\bar{\mathbf{M}}$ ($\equiv \mathbf{P}_{\bar{\boldsymbol{\theta}}}$) represents orthogonal projection along $\bar{\boldsymbol{\theta}}$. We have also noted that equating the first order terms in the expansion of $\|\bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots\|^2 = 1$ results in $(\bar{\boldsymbol{\theta}}, \Delta_1\boldsymbol{\theta}) = 0$ (Kanatani, 2008), hence $\mathbf{P}_{\bar{\boldsymbol{\theta}}}\Delta_1\boldsymbol{\theta} = \Delta_1\boldsymbol{\theta}$. Substituting (31) into (30), we can express $\Delta_2\lambda$ in the form

$$\Delta_2\lambda = \frac{(\bar{\boldsymbol{\theta}}, \Delta_2\mathbf{M}\bar{\boldsymbol{\theta}}) - (\bar{\boldsymbol{\theta}}, \Delta_1\mathbf{M}\bar{\mathbf{M}}^-\Delta_1\mathbf{M}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \mathbf{N}\bar{\boldsymbol{\theta}})} = \frac{(\bar{\boldsymbol{\theta}}, \mathbf{T}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \mathbf{N}\bar{\boldsymbol{\theta}})}, \tag{32}$$

where

$$\mathbf{T} = \Delta_2\mathbf{M} - \Delta_1\mathbf{M}\bar{\mathbf{M}}^-\Delta_1\mathbf{M}. \tag{33}$$

Next, we consider the second order error $\Delta_2\boldsymbol{\theta}$. Since $\bar{\boldsymbol{\theta}}$ is a unit vector and does not change its norm, we are interested in the error component orthogonal to $\bar{\boldsymbol{\theta}}$ (the first order error $\Delta_1\boldsymbol{\theta}$ in (31) is orthogonal to $\bar{\boldsymbol{\theta}}$ as is). We define the orthogonal component of $\Delta_2\boldsymbol{\theta}$ by

$$\Delta_2^\perp\boldsymbol{\theta} = \mathbf{P}_{\bar{\boldsymbol{\theta}}}\Delta_2\boldsymbol{\theta} \ (= \bar{\mathbf{M}}^-\bar{\mathbf{M}}\Delta_2\boldsymbol{\theta}). \tag{34}$$

Multiplying (30) by $\bar{\mathbf{M}}^-$ from left and substituting (31), we obtain

$$\Delta_2^\perp\boldsymbol{\theta} = \Delta_2\lambda\bar{\mathbf{M}}^-\mathbf{N}\bar{\boldsymbol{\theta}} + \bar{\mathbf{M}}^-\Delta_1\mathbf{M}\bar{\mathbf{M}}^-\Delta_1\mathbf{M}\bar{\boldsymbol{\theta}} - \bar{\mathbf{M}}^-\Delta_2\mathbf{M}\bar{\boldsymbol{\theta}} = \frac{(\bar{\boldsymbol{\theta}}, \mathbf{T}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \mathbf{N}\bar{\boldsymbol{\theta}})}\bar{\mathbf{M}}^-\mathbf{N}\bar{\boldsymbol{\theta}} - \bar{\mathbf{M}}^-\mathbf{T}\bar{\boldsymbol{\theta}}. \tag{35}$$

4. Covariance and bias

4.1. *General algebraic fitting*

From (31), we see that the leading term of the covariance matrix of the solution $\boldsymbol{\theta}$ is given by

$$\begin{aligned} V[\boldsymbol{\theta}] &= E[\Delta_1\boldsymbol{\theta}\Delta_1\boldsymbol{\theta}^\top] = \bar{\mathbf{M}}^-E[(\Delta_1\mathbf{M}\boldsymbol{\theta})(\Delta_1\mathbf{M}\boldsymbol{\theta})^\top]\bar{\mathbf{M}}^- = \frac{1}{N^2}\bar{\mathbf{M}}^-E\Big[\sum_{\alpha=1}^N(\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{\theta})\bar{\boldsymbol{\xi}}_\alpha\sum_{\beta=1}^N(\Delta\boldsymbol{\xi}_\beta, \boldsymbol{\theta})\bar{\boldsymbol{\xi}}_\beta^\top\Big]\bar{\mathbf{M}}^- \\ &= \frac{1}{N^2}\bar{\mathbf{M}}^-\sum_{\alpha,\beta=1}^N(\boldsymbol{\theta}, E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{\theta})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top\bar{\mathbf{M}}^- = \frac{\sigma^2}{N^2}\bar{\mathbf{M}}^-\Big(\sum_{\alpha=1}^N(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\Big)\bar{\mathbf{M}}^- = \frac{\sigma^2}{N}\bar{\mathbf{M}}^-\bar{\mathbf{M}}'\bar{\mathbf{M}}^-, \end{aligned} \tag{36}$$

where we define

$$\bar{\mathbf{M}}' = \frac{1}{N}\sum_{\alpha=1}^N(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top. \tag{37}$$

In deriving (36), we have noted that $\boldsymbol{\xi}_\alpha$ is independent for different α and that $E[\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\beta^\top] = \delta_{\alpha\beta}\sigma^2V_0[\boldsymbol{\xi}_\alpha]$, where $\delta_{\alpha\beta}$ is the Kronecker delta. The important observation is that the covariance matrix $V[\boldsymbol{\theta}]$, which is $O(\sigma^2)$, does not depend on $\mathbf{N}$. This means that *all algebraic methods have the same the covariance matrix in the leading order*, as pointed out by Al-Sharadqah and Chernov (2009) for circle fitting. Thus, we are unable to reduce the covariance of $\boldsymbol{\theta}$ by adjusting $\mathbf{N}$. This may sound contradictory to the fact that the method of Taubin (1991), an algebraic method using $\mathbf{N}_\mathrm{T}$ in (23), is known to be far more accurate than the standard LS (Kanatani and Sugaya, 2007). We now show that the accuracy difference stems from the *bias* terms and that a better method can be obtained by further reducing the bias.

Since $E[\Delta_1\boldsymbol{\theta}] = \mathbf{0}$, there is no bias in the first order; the leading bias is $O(\sigma^2)$. In order to evaluate the second order bias $E[\Delta_2^\perp\boldsymbol{\theta}]$, we need to evaluate the expectation of $\mathbf{T}$ in (33). To do so, we first consider the term $E[\Delta_2\mathbf{M}]$. From (20) and (21), we see that

$$E[\Delta_2\boldsymbol{\xi}_\alpha] = \sigma^2\mathbf{e}, \qquad \mathbf{e} \equiv \begin{cases} (2,0,0,0)^\top & \text{for circles} \\ (1,0,1,0,0,0)^\top & \text{for ellipses} \end{cases}. \tag{38}$$

Hence, we see from the second equation in (25) that

$$E[\Delta_2\mathbf{M}] = \frac{1}{N}\sum_{\alpha=1}^{N}\Big(\bar{\boldsymbol{\xi}}_\alpha E[\Delta_2\boldsymbol{\xi}_\alpha]^\top + E[\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top] + E[\Delta_2\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\xi}}_\alpha^\top\Big) = \sigma^2\Big(\mathbf{N}_{\mathrm{T}} + 2\mathcal{S}[\bar{\boldsymbol{\xi}}_c\mathbf{e}^\top]\Big), \tag{39}$$

where we have noted $E[\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top] = \sigma^2 V_0[\boldsymbol{\xi}_\alpha]$ and used (23). The symbol $\mathcal{S}[\cdot]$ denotes symmetrization ($\mathcal{S}[\mathbf{A}] \equiv (\mathbf{A} + \mathbf{A}^\top)/2$). The vector $\bar{\boldsymbol{\xi}}_c$ is defined by

$$\bar{\boldsymbol{\xi}}_c = \frac{1}{N}\sum_{\alpha=1}^{N}\bar{\boldsymbol{\xi}}_\alpha. \tag{40}$$

We next consider the term $E[\Delta_1\mathbf{M}\bar{\mathbf{M}}^-\Delta_1\mathbf{M}]$. It has the form

$$E[\Delta_1\mathbf{M}\bar{\mathbf{M}}^-\Delta_1\mathbf{M}] = \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N}\Big(\mathrm{tr}[\bar{\mathbf{M}}^- V_0[\boldsymbol{\xi}_\alpha]]\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top + (\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[V_0[\boldsymbol{\xi}_\alpha]\bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top]\Big), \tag{41}$$

where $\mathrm{tr}[\cdot]$ denotes the trace (see Appendix B for the derivation). From (39) and (41), the matrix $\mathbf{T}$ in (33) has the following expectation:

$$E[\mathbf{T}] = \sigma^2\Big(\mathbf{N}_{\mathrm{T}} + 2\mathcal{S}[\bar{\boldsymbol{\xi}}_c\mathbf{e}^\top] - \frac{1}{N^2}\sum_{\alpha=1}^{N}\Big(\mathrm{tr}[\bar{\mathbf{M}}^- V_0[\boldsymbol{\xi}_\alpha]]\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top + (\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[V_0[\boldsymbol{\xi}_\alpha]\bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top]\Big)\Big). \tag{42}$$

Thus, the second order error $\Delta_2^\perp\boldsymbol{\theta}$ in (35) has the following bias:

$$E[\Delta_2^\perp\boldsymbol{\theta}] = \bar{\mathbf{M}}^-\Big(\frac{(\bar{\boldsymbol{\theta}}, E[\mathbf{T}]\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \mathbf{N}\bar{\boldsymbol{\theta}})}\mathbf{N}\bar{\boldsymbol{\theta}} - E[\mathbf{T}]\bar{\boldsymbol{\theta}}\Big). \tag{43}$$

4.2. *Standard LS*

From (42), we see that $(\bar{\boldsymbol{\xi}}_c, \bar{\boldsymbol{\theta}}) = \mathbf{0}$ and $(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{\theta}}) = \mathbf{0}$. Hence, $E[\mathbf{T}]\bar{\boldsymbol{\theta}}$ can be written as

$$E[\mathbf{T}]\bar{\boldsymbol{\theta}} = \sigma^2\Big(\mathbf{N}_{\mathrm{T}}\bar{\boldsymbol{\theta}} + (\mathbf{e}, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_c - \frac{1}{N^2}\sum_{\alpha=1}^{N}\Big((\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} + (\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha\Big)\Big). \tag{44}$$

For standard LS ($\mathbf{N} = \mathbf{I}$), we see from (43) that

$$E[\Delta_2^\perp\boldsymbol{\theta}] = \bar{\mathbf{M}}^-\Big((\bar{\boldsymbol{\theta}}, E[\mathbf{T}]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\theta}} - E[\mathbf{T}]\bar{\boldsymbol{\theta}}\Big) = -\bar{\mathbf{M}}^-(\mathbf{I} - \bar{\boldsymbol{\theta}}\bar{\boldsymbol{\theta}}^\top)E[\mathbf{T}]\bar{\boldsymbol{\theta}} = -\bar{\mathbf{M}}^- E[\mathbf{T}]\bar{\boldsymbol{\theta}}, \tag{45}$$

where we have used the following equality:

$$\bar{\mathbf{M}}^-(\mathbf{I} - \bar{\boldsymbol{\theta}}\bar{\boldsymbol{\theta}}^\top) = \bar{\mathbf{M}}^-\mathbf{P}_{\bar{\boldsymbol{\theta}}} = \bar{\mathbf{M}}^-\bar{\mathbf{M}}\bar{\mathbf{M}}^- = \bar{\mathbf{M}}^-. \tag{46}$$

From (44), and (45), the leading bias of the standard LS has the following form:

$$E[\Delta_2^\perp\boldsymbol{\theta}] = -\sigma^2\bar{\mathbf{M}}^-\Big(\mathbf{N}_{\mathrm{T}}\bar{\boldsymbol{\theta}} + (\mathbf{e}, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_c - \frac{1}{N^2}\sum_{\alpha=1}^{N}\Big((\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} + (\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha\Big)\Big). \tag{47}$$

4.3. *Taubin method*

If we note that $(\bar{\boldsymbol{\xi}}_c, \bar{\boldsymbol{\theta}}) = \mathbf{0}$ and $(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{\theta}}) = \mathbf{0}$, we see from (42) that

$$\begin{aligned}(\bar{\boldsymbol{\theta}}, E[\mathbf{T}]\bar{\boldsymbol{\theta}}) &= \sigma^2\Big((\bar{\boldsymbol{\theta}}, \mathbf{N}_{\mathrm{T}}\bar{\boldsymbol{\theta}}) - \frac{1}{N^2}\sum_{\alpha=1}^{N}(\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha)(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\Big) \\ &= \sigma^2\Big((\bar{\boldsymbol{\theta}}, \mathbf{N}_{\mathrm{T}}\bar{\boldsymbol{\theta}}) - \frac{1}{N^2}\sum_{\alpha=1}^{N}\mathrm{tr}[\bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top](\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\Big) \\ &= \sigma^2\Big((\bar{\boldsymbol{\theta}}, \mathbf{N}_{\mathrm{T}}\bar{\boldsymbol{\theta}}) - \frac{1}{N^2}\mathrm{tr}[\bar{\mathbf{M}}^-\sum_{\alpha=1}^{N}(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top]\Big) = \sigma^2(\bar{\boldsymbol{\theta}}, \mathbf{N}_{\mathrm{T}}\bar{\boldsymbol{\theta}}) - \frac{\sigma^2}{N}\mathrm{tr}[\bar{\mathbf{M}}^-\bar{\mathbf{M}}'],\end{aligned} \tag{48}$$

where we have used (37). For the Taubin method ($\mathbf{N} = \mathbf{N}_{\mathrm{T}}$), the leading bias (43) is

$$E[\Delta_2^{\perp}\boldsymbol{\theta}] = -\sigma^2\bar{\mathbf{M}}^{-}\Big(q\mathbf{N}_{\mathrm{T}}\bar{\boldsymbol{\theta}} + (\mathbf{e},\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_c - \frac{1}{N^2}\sum_{\alpha=1}^{N}\Big((\bar{\boldsymbol{\xi}}_\alpha,\bar{\mathbf{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} + (\bar{\boldsymbol{\theta}},V_0[\boldsymbol{\xi}_\alpha]\bar{\mathbf{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha\Big)\Big), \tag{49}$$

where we put

$$q = \frac{1}{N}\frac{\mathrm{tr}[\bar{\mathbf{M}}^{-}\bar{\mathbf{M}}']}{(\bar{\boldsymbol{\theta}},\mathbf{N}_{\mathrm{T}}\bar{\boldsymbol{\theta}})}. \tag{50}$$

Comparing (49) and (47), we notice that the only difference is that $\mathbf{N}_{\mathrm{T}}\bar{\boldsymbol{\theta}}$ in (47) is replaced by $q\mathbf{N}_{\mathrm{T}}\bar{\boldsymbol{\theta}}$ in (49). We see from (50) that $q < 1$ when N is large. This can be regarded as one of the reasons of the high accuracy of the Taubin method (Kanatani, 2008).

5. Hyper least squares

Now, we propose to chose the matrix $\mathbf{N}$ to be

$$\mathbf{N} = \mathbf{N}_{\mathrm{T}} + 2\mathcal{S}[\bar{\boldsymbol{\xi}}_c\mathbf{e}^{\top}] - \frac{1}{N^2}\sum_{\alpha=1}^{N}\Big(\mathrm{tr}[\bar{\mathbf{M}}^{-}V_0[\boldsymbol{\xi}_\alpha]]\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^{\top} + (\bar{\boldsymbol{\xi}}_\alpha,\bar{\mathbf{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[V_0[\boldsymbol{\xi}_\alpha]\bar{\mathbf{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^{\top}]\Big). \tag{51}$$

Then, we have $E[\mathbf{T}] = \sigma^2\mathbf{N}$ from (42), so (43) becomes

$$E[\Delta_2^{\perp}\boldsymbol{\theta}] = \sigma^2\bar{\mathbf{M}}^{-}\Big(\frac{(\bar{\boldsymbol{\theta}},\mathbf{N}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}},\mathbf{N}\bar{\boldsymbol{\theta}})}\mathbf{N} - \mathbf{N}\Big)\bar{\boldsymbol{\theta}} = \mathbf{0}. \tag{52}$$

Thus, the choice of $\mathbf{N}$ in (51) completely eliminates the bias up to second order noise terms. However, (51) involves the true values $\bar{\boldsymbol{\xi}}_\alpha$ and $\bar{\mathbf{M}}$, so we evaluate them by replacing the true values $(\bar{x}_\alpha,\bar{y}_\alpha)$ in their definitions by the observations (x_α,y_α). As a result, the matrix $\bar{\mathbf{M}}$ becomes nonsingular, so we compute its spectral decomposition, replace the smallest eigenvalue with 0, and evaluate the pseudoinverse $\bar{\mathbf{M}}^{-}$. The use of observations in the place of their true values does not affect the order analysis, because expectations of odd-order error terms vanish and hence the error in (52) is at most $O(\sigma^4)$. Thus, *the second order bias is exactly 0*. Following Al-Sharadqah and Chernov (2009), we call the method using (51) *hyper least squares*, or *HyperLS*.

Since the last term on the right hand side of (51) is $O(1/N)$, we may omit it. Let us call the method using

$$\mathbf{N}_{\mathrm{S}} = \mathbf{N}_{\mathrm{T}} + 2\mathcal{S}[\bar{\boldsymbol{\xi}}_c\mathbf{e}^{\top}], \tag{53}$$

SemihyperLS. It seems that the accuracy is preserved, since N is usually large. According to our experiments reported in the next section, however, this is not necessarily so when N is not very large. Now, we summarize the resulting schemes, replacing the true values by observations. For circles, we have

$$\boldsymbol{\xi}_c = (s_x^2+s_y^2, f_0x_c, f_0y_c, f_0^2)^{\top}, \qquad \mathbf{e} = (2,0,0,0)^{\top}, \tag{54}$$

where we define

$$x_c = \frac{1}{N}\sum_{\alpha=1}^{N}x_\alpha, \quad y_c = \frac{1}{N}\sum_{\alpha=1}^{N}y_\alpha, \quad s_x^2 = \frac{1}{N}\sum_{\alpha=1}^{N}x_\alpha^2, \quad s_y^2 = \frac{1}{N}\sum_{\alpha=1}^{N}y_\alpha^2. \tag{55}$$

The Taubin method and the SemihyperLS, respectively, correspond to

$$\mathbf{N}_{\mathrm{T}} = \begin{pmatrix} 4(s_x^2+s_y^2) & 2f_0x_c & 2f_0y_c & 0 \\ 2f_0x_c & f_0^2 & 0 & 0 \\ 2f_0y_c & 0 & f_0^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{N}_{\mathrm{S}} = \begin{pmatrix} 8(s_x^2+s_y^2) & 4f_0x_c & 4f_0y_c & 2f_0^2 \\ 4f_0x_c & f_0^2 & 0 & 0 \\ 4f_0y_c & 0 & f_0^2 & 0 \\ 2f_0^2 & 0 & 0 & 0 \end{pmatrix}. \tag{56}$$

The use of the above $\mathbf{N}_{\mathrm{S}}$ is the circle fitting algorithm of Al-Sharadqah and Chernov (2009), who called it "Hyper". For ellipses, we have

$$\boldsymbol{\xi}_c = (s_x^2, 2\gamma_{xy}, s_y^2, 2f_0x_c, 2f_0y_c, f_0^2)^{\top}, \qquad \mathbf{e} = (1,0,1,0,0,0)^{\top}, \tag{57}$$

where we define

$$\gamma_{xy} = \frac{1}{N}\sum_{\alpha=1}^{N}x_\alpha y_\alpha. \tag{58}$$

The Taubin method and the SemihyperLS, respectively, correspond to

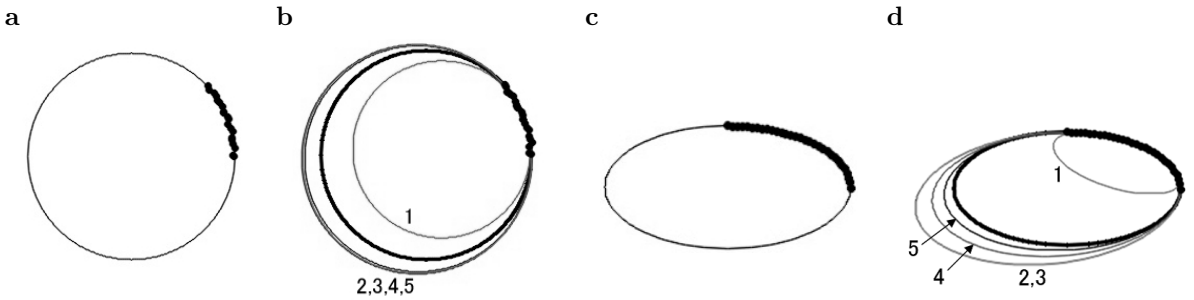

Fig. 1. **a**. Twenty points on a circle. **b**. Fitted circles for $\sigma = 1.0$. **c**. Thirty points on an ellipse. **d**. Fitted ellipses for $\sigma = 0.5$. In **b** and **d**, the lines correspond to 1. standard LS, 2. Taubin method, 3. SemihyperLS, 4. HyperLS, 5. ML. The true shape is drawn in thick lines.

$$\mathbf{N}_{\mathrm{T}}=4\begin{pmatrix} s_x^2 & \gamma_{xy} & 0 & f_0x_c & 0 & 0\\ \gamma_{xy} & s_x^2+s_y^2 & \gamma_{xy} & f_0y_c & f_0x_c & 0\\ 0 & \gamma_{xy} & s_y^2 & 0 & f_0y_c & 0\\ f_0x_c & f_0y_c & 0 & f_0^2 & 0 & 0\\ 0 & f_0x_c & f_0y_c & 0 & f_0^2 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{N}_{\mathrm{S}}=\begin{pmatrix} 6s_x^2 & 6\gamma_{xy} & s_x^2+s_y^2 & 6f_0x_c & 2f_0y_c & f_0^2\\ 6\gamma_{xy} & 4(s_x^2+s_y^2) & 6\gamma_{xy} & 4f_0y_c & 4f_0x_c & 0\\ s_x^2+s_y^2 & 6\gamma_{xy} & 6s_y^2 & 2f_0x_c & 6f_0y_c & f_0^2\\ 6f_0x_c & 4f_0y_c & 2f_0x_c & 4f_0^2 & 0 & 0\\ 2f_0y_c & 4f_0x_c & 6f_0y_c & 0 & 4f_0^2 & 0\\ f_0^2 & 0 & f_0^2 & 0 & 0 & 0 \end{pmatrix}. \tag{59}$$

6. Numerical experiments

We placed 20 equidistant points subtending $\pi/4$ on a circle of radius 100 (Fig. 1**a**) and 30 equidistant points in the first quadrant of an ellipse of major and minor axes 100 and 50 (Fig. 1**c**). The scale constant f_0 is fixed to 600, meaning that we are in effect considering a circle of radius 0.17 and an ellipse of major and minor axes 0.17 and 0.08 in an image of size 1×1. We added to the x- and y-coordinates of each point independent Gaussian noise of mean 0 and standard deviation σ and fitted a circle and an ellipse by the standard LS, the Taubin method, the SemihyperLS, the (full) HyperLS, and ML, which minimizes the geometric distance

$$J_{\mathrm{ML}} = \frac{1}{N}\sum_{\alpha=1}^{N}\Big((x_\alpha-\bar{x}_\alpha)^2+(y_\alpha-\bar{y}_\alpha)^2\Big), \tag{60}$$

rather than the algebraic distance of (14), subject to the constraint that each $(\bar{x}_\alpha, \bar{y}_\alpha)$ satisfies the circle/ellipse equation in (1) or (2). This maximizes the likelihood if the noise is independent, identical, and isotropic Gaussian. For numerical computation, we used the FNS of Chojnacki et al. (2000) (see Kanatani and Sugaya (2007) for the details). As mentioned in the Introduction, FNS and similar schemes like HEIV and projective Gauss-Newton iterations minimize not directly (60) but a simplified form called the Sampson error, but the minimum of (60) can be obtained by running a few iterations of a Sampson error minimization algorithm (Kanatani and Sugaya, 2010). It has also been observed that the solution that minimizes the Sampson error agrees with the ML solution up to several significant digits (Kanatani and Sugaya, 2008). Thus, FNS can safely be regarded as minimizing (60).

Figure 1**b** show fitted circles for $\sigma = 1.0$, and Fig. 1**d** ellipses for $\sigma = 0.5$. As we see, circle fitting is very robust to noise; except for the standard LS, which is very poor, not much difference is visible among different methods even for large noise. In contrast, ellipse fitting is very sensitive to noise, causing large differences among methods even for small noise. For statistical evaluation, we conducted the following experiments.

Since the computed $\boldsymbol{\theta}$ and its true value $\bar{\boldsymbol{\theta}}$ are both unit vectors, we measure their discrepancy by the orthogonal component

$$\Delta^{\perp}\boldsymbol{\theta} = \mathbf{P}_{\bar{\boldsymbol{\theta}}}\boldsymbol{\theta}, \tag{61}$$

where $\mathbf{P}_{\bar{\boldsymbol{\theta}}}$ ($\equiv \mathbf{I} - \bar{\boldsymbol{\theta}}\bar{\boldsymbol{\theta}}^{\top}$) is the orthogonal projection matrix along $\bar{\boldsymbol{\theta}}$ (Fig. 2). We generated 10000 independent noise instances for each σ and evaluated the bias B and the RMS (root-mean-square) error D defined by

$$B = \Big\|\frac{1}{10000}\sum_{a=1}^{10000}\Delta^{\perp}\boldsymbol{\theta}^{(a)}\Big\|, \qquad D = \sqrt{\frac{1}{10000}\sum_{a=1}^{10000}\|\Delta^{\perp}\boldsymbol{\theta}^{(a)}\|^2}, \tag{62}$$

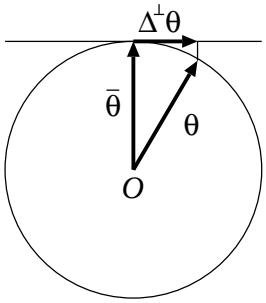

Fig. 2. The true value $\bar{\boldsymbol{\theta}}$, the computed value $\boldsymbol{\theta}$, and its orthogonal component $\Delta^{\perp}\boldsymbol{\theta}$ to $\bar{\boldsymbol{\theta}}$.

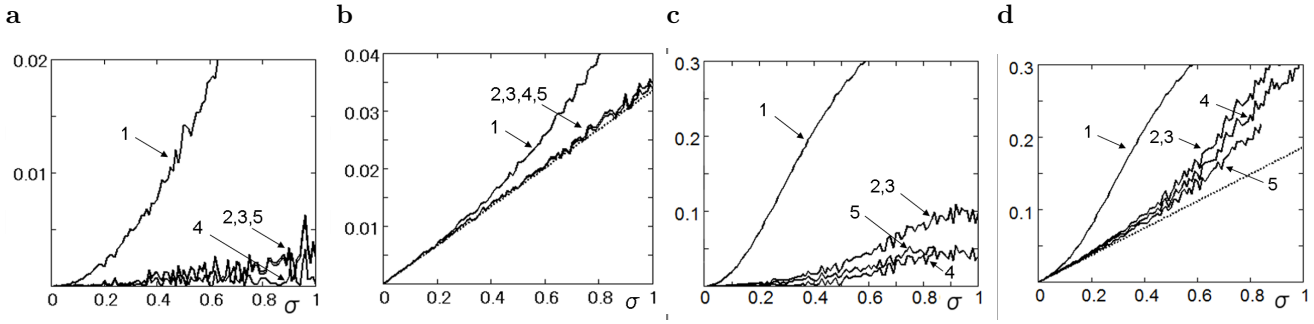

Fig. 3. The bias (**a**) and the RMS error (**b**) for circle fitting, and the bias (**c**) and the RMS error (**d**) for ellipse fitting. The horizontal axes indicate the standard deviation σ of the added noise. 1. Standard LS. 2. Taubin method. 3. SemihyperLS. 4. HyperLS. 5. ML. The dotted lines in **b** and **d** indicate the KCR lower bound.

where $\boldsymbol{\theta}^{(a)}$ is the solution in the ath trial. Figure 3 plots of B and D for the circle (**a**, **b**) and the ellipse (**c**, **d**). The dotted lines in **b** and **d** indicate the theoretical limit, called the *KCR lower bound* (Chernov and Lesort, 2004; Kanatani, 1996, 2008), given by

$$D_{\rm KCR} = \sigma\sqrt{\mathrm{tr}[\Big(\sum_{\alpha=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})}\Big)^{-}]}. \tag{63}$$

A similar expression was also obtained by Ameiya and Fuller (1988). It is known that the ML solution reaches this bound in the leading order (Chernov and Lesort, 2004; Kanatani, 1996, 2008).

Standard linear algebra routines for solving the generalized eigenvalue problem in the form of (17) assumes that the matrix $\mathbf{N}$ is positive definite. As can be seen from (56) and (59), however, the matrix $\mathbf{N}_{\rm T}$ has a row and a column of zeros. It is easy to see that the matrix $\mathbf{N}_{\rm S}$ in (56) and (59) is not positive definite, and numerical tests show that the matrix $\mathbf{N}$ in (51) is not necessarily positive definite, either. However, this causes no problem, because (17) can be written as

$$\mathbf{N}\boldsymbol{\theta} = \frac{1}{\lambda}\mathbf{M}\boldsymbol{\theta}. \tag{64}$$

Since the matrix $\mathbf{M}$ in (15) is positive definite for noisy data, we can solve (64) instead, using a standard routine. If the smallest eigenvalue of $\mathbf{M}$ happens to be 0, it indicates that the data are all exact; any method, e.g., the standard LS, gives an exact solution. For noisy data, the solution $\boldsymbol{\theta}$ is given by the eigenvector of (64) for the eigenvalue $1/\lambda$ with the largest absolute value.

As we can see from Fig. 3**a**, **b**, there do exist accuracy differences among different methods of circles, but they are very small except the standard LS. However, Fig. 3**a** shows that HyperLS has smaller bias than all other methods. For ellipses, the differences among methods are more marked. Fig. 3**a**, **c** shows that for both circles and ellipses, the standard LS has very large bias, as compared to which the Taubin solution has much smaller bias, and the HyperLS has even smaller bias. The bias of SemihyperLS is nearly the same as the Taubin method for both circles and ellipses. As pointed out in Section 4, all algebraic methods have the same covariance matrix to the leading order, meaning that the accuracy difference among algebraic methods are mainly accounted for by the bias. This is confirmed by Fig. 3**b**, **d**. For both circles and ellipses, the standard LS performs very poorly, compared to which the Taubin solution is significantly better when measured in the RMS error. For ellipses, the HyperLS is more accurate than the Taubin method. The SemihyperLS performs nearly like the Taubin method for both circles and ellipses.

Figure 3**a**, **c** shows that the bias of HyperLS is even smaller than ML, nearly zero for small σ as theoretically predicted. Yet, Fig. 3**d** shows that ML has the smallest RMS error of all. This is because ML minimizes the geometric distance of

(60), rather than the algebraic distance, and hence has a smaller covariance matrix (Kanatani, 1996, 2008). However, ML computation may not converge in the presence of large noise. Indeed, the interrupted plot of ML in Fig. 3**d** indicate that the iterations did not converge beyond that noise level. For circles, ML is very stable but also fails for very large noise. In the example shown here, it failed around the noise level $\sigma = 4$. The convergence of ML critically depends on the accuracy of the initialization. Here, we used the standard LS to start the ML iterations. We confirmed that the use of our HyperLS to start ML significantly extends the noise range of convergence, though the computation fails sooner or later.

7. Conclusions

We have successfully extended the HyperLS principle, originally invented for circle fitting (Rangarajan and Kanatani, 2009), to ellipses [2] . This is achieved by finding a scale normalization that eliminates the statistical bias of the LS estimator up to second order noise terms. To this end, we performed a rigorous statistical analysis of the covariance and bias of the LS ellipse fitting problem. Doing numerical experiments, we validated our findings by comparing the performance of our estimator with other LS estimators. Since the results in Secion 6 are limited to particular instances, and the observed differences are rather small expect for the standard LS, they may not be sufficient to draw definitive conclusions. However, additional experiments suggest that:

(i) For fitting circles/ellipses to points that span more than 1/4 of the circumference, all LS estimators have nearly identical performance; even the standard LS produces a fairly accurate solution (see Kanatani and Sugaya (2007) for comparative experiments).

(ii) For short sequences, the behavior of the standard LS is unpredictable. Most often, it performs very poorly, producing a very small circle/ellipse, as shown in our example, but sometimes it produces a fairly good fit, depending on the shape of the sequence and data scaling or the choice of f_0. It thus is more appropriate to say that the standard LS is "unreliable" than "inaccurate". We infer that the strong bias of the standard LS may sometimes favorably shift the solution toward the true shape.

(iii) The differences among Taubin, SemihyperLS, HyperLS, and ML are generally very small. What we can say with certainty is that Taubin and SemihyperLS almost always produce practically the same fits. In comparison, HyperLS fits are generally better, although the difference is small and difficult to perceive by merely observing particular examples. The distinction can be observed only after statistical tests using a lot of different noise instances. In our experiments, we averaged over 10,000 trials, but this still appears insufficient to approximate the expectation, as evidenced by the large variations for different σ.

(iv) ML produces almost always the best fit, but the failure of convergence in the presence of large noise is unavoidable. The convergence heavily depends on the accuracy of the initialization.

(v) Comparing Taubin, SemihyperLS, and HyperLS, we could not find any evidence that Taubin or SemihyperLS should be preferred to HyperLS. We also observed that the use of HyperLS for ML initialization can significantly extend the range of convergence. We conclude that initializing ML iterations with the proposed HyperLS is the best choice for circle/ellipse fitting.

In this paper, we analyzed only the leading covariance term, which is $O(\sigma^2)$, and the leading bias term, which is also $O(\sigma^2)$. Due to technical difficulties, we are unable to evaluate at this stage how higher order terms of $O(\sigma^4)$ affect the solution. This is a remaining issue to be studied in the future.

Acknowledgments

The authors thank Yuuki Iwamoto of Okayama University for his assistance in numerical experiments. We also thank Ali Al-Sharadqah and Nikolai Chernov of the University of Alabama at Birmingham, U.S.A, Wolfgang Förstner of the University of Bonn, Germany, and Alexander Kukush of National Taras Shevchenko University of Kyiv, Ukraine, for helpful discussions on this research. This work was supported in part by the Ministry of Education, Culture, Sports, Science, and Technology, Japan, under a Grant in Aid for Scientific Research (C 21500172).

Appendix A. Scale effects of algebraic solutions

Motivation. In (1) and (2), f_0 is a scale constant of the order of the magnitude of the data. This in effect scales the x and y coordinates of the data into the range ± 1. This f_0 is introduced to stabilize numerical computation with finite precision length; without it, serious accuracy loss is incurred when we use real image data in image processing applications. Suppose, for example, the image size is about 1000×1000 pixels, and suppose we are considering a circle of radius of about 100 pixels. If we let $f_0 = 1$, the vector $\boldsymbol{\xi}$ in (11) is about $(100^2 + 100^2, ..., 1)^\top = (20000, ..., 1)^\top$. For its norm $\|\boldsymbol{\xi}\|$, we compute $\sqrt{400000000 + \cdots + 1}$, but the value 1 would easily be lost in the course of finite length computation, resulting

[2] The code is available at http://homepages.inf.ed.ac.uk/cgi/rbf/CVONLINE/entries.pl?TAG384.

in a significant loss of accuracy. Here, we briefly discuss, purely for *theoretical interest*, how the solution of (17) would be affected if we use different values of f_0, assuming the computation is done using a hypothetical computer with infinite precision length. We should keep in mind that this analysis has no practical meaning in real situations.

Theory. Let us consider the case of ellipse fitting. If we magnify the data coordinates (x, y) s times to (sx, sy), the vector $\boldsymbol{\xi}$ in (12) is transformed to $\boldsymbol{\xi}' = \mathbf{S}\boldsymbol{\xi}$, where $\mathbf{S} = \mathrm{diag}(s^2, s^2, s^2, s, s, 1)$. If an ellipse having coefficients $\boldsymbol{\theta} = (A, B, C, D, E, F)^\top$ is enlarged by s times, the enlarged ellipse has coefficients $\boldsymbol{\theta}' = (A/s^2, B/s^2, C/s^2, D/s, E/s, F)^\top$, so $\boldsymbol{\theta} = \mathbf{S}\boldsymbol{\theta}'$ holds. By this scale change, the matrix $\mathbf{M}$ in (15) is transformed to $\mathbf{M}' = \mathbf{SMS}$. Let $\mathbf{M}'\boldsymbol{\theta}' = \lambda'\mathbf{N}'\boldsymbol{\theta}'$ be the transformed eigenvalue problem of $\mathbf{M}\boldsymbol{\theta} = \lambda\mathbf{N}\boldsymbol{\theta}$. It is easy to see that if $\mathbf{N}' = \beta\mathbf{SNS}$ for some $\beta \neq 0$, the transformed problem has the solution $\boldsymbol{\theta}' = \mathbf{S}^{-1}\boldsymbol{\theta}$ with $\lambda' = \beta\lambda$. Hence, the ellipse fitted to the data enlarged s times is the s times enlargement of the ellipse fitted to the original data. If this is the case, we say that the fitting method is *scale invariant*.

Examples. It is easy to see that $\mathbf{N} = \beta\mathbf{SNS}$ holds for the matrices $\mathbf{N}$ corresponding to (5), (6), (9), and (10), so the methods using them are scale invariant. However, the use of (7) and (8) is not scale invariant. We can easily see that the matrix $\mathbf{N}_{\mathrm{T}}$ in (56) transforms to $\mathbf{N}'_{\mathrm{T}} = \mathbf{SN}_{\mathrm{T}}\mathbf{S}/s^2$, so the Taubin method is also scale invariant. However, this must *not* be taken to mean that the Taubin method computes the same ellipse enlarged/reduced if the data are enlarged/reduced, because computer computation does not return a correct value unless carried out with an appropriate scale f_0.

HyperLS. We can easily see that the matrix $\mathbf{N}_S$ in (56) transforms to $\mathbf{N}'_{\mathrm{S}} = \mathbf{SN}_{\mathrm{S}}\mathbf{S}/s^2$, so SemihyperLS is also scale invariant. However, HyperLS is not strictly invariant because it involves the pseudoinverse $\bar{\mathbf{M}}^-$. Since $\bar{\mathbf{M}}' = \mathbf{S}\bar{\mathbf{M}}\mathbf{S}$, we would obtain $\bar{\mathbf{M}}'^{-1} = \mathbf{S}^{-1}\bar{\mathbf{M}}^{-1}\mathbf{S}^{-1}$ if $\bar{\mathbf{M}}$ were nonsingular. However, $\bar{\mathbf{M}}'^{-} = \mathbf{S}^{-1}\bar{\mathbf{M}}^{-}\mathbf{S}^{-1}$ does not necessarily hold for pseudoinverse. If this were to hold, we can easily see that the matrix $\mathbf{N}$ in (51) would satisfy to $\mathbf{N}' = \mathbf{SNS}/s^2$, guaranteeing the scale invariance of HyperLS. Geometrically, the pseudoinverse is an operation of decomposing the entire space into the direct sum of the null space of the matrix and its orthogonal complement and defining the inverse only in the orthogonal complement keeping the null space intact. If the space is linearly transformed by $\mathbf{S}$, the null space is also transformed accordingly, and $\bar{\mathbf{M}}'^{-} = \mathbf{S}^{-1}\bar{\mathbf{M}}^{-}\mathbf{S}^{-1}$ no longer holds.

Remarks. How much is the HyperLS solution affected by the value of f_0 if its variations are kept *within the order of the magnitude of the data* (otherwise, correct computation cannot be done by a finite length computer)? This can be checked only by experiments. According to our experiments, $\boldsymbol{\theta}$ and $\mathbf{S}\boldsymbol{\theta}'$, after normalized into unit vectors, share several significant digits, as long as the scale change is kept to the order of 1 – 10. In contrast, the difference is much larger for the standard LS. Thus, we can regard HyperLS as *practically* scale invariant.

Appendix B. Derivation of (41)

$$
\begin{aligned}
E[\Delta_1\mathbf{M}\bar{\mathbf{M}}^-\Delta_1\mathbf{M}] &= E[\frac{1}{N}\sum_{\alpha=1}^{N}\Big(\bar{\boldsymbol{\xi}}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top + \Delta_1\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\Big)\bar{\mathbf{M}}^-\frac{1}{N}\sum_{\beta=1}^{N}\Big(\bar{\boldsymbol{\xi}}_\beta\Delta_1\boldsymbol{\xi}_\beta^\top + \Delta_1\boldsymbol{\xi}_\beta\bar{\boldsymbol{\xi}}_\beta^\top\Big)] \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} E[(\bar{\boldsymbol{\xi}}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top + \Delta_1\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top)\bar{\mathbf{M}}^-(\bar{\boldsymbol{\xi}}_\beta\Delta_1\boldsymbol{\xi}_\beta^\top + \Delta_1\boldsymbol{\xi}_\beta\bar{\boldsymbol{\xi}}_\beta^\top) \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} E[\bar{\boldsymbol{\xi}}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top\bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\beta\Delta_1\boldsymbol{\xi}_\beta^\top + \bar{\boldsymbol{\xi}}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top\bar{\mathbf{M}}^-\Delta_1\boldsymbol{\xi}_\beta\bar{\boldsymbol{\xi}}_\beta^\top + \Delta_1\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\beta\Delta_1\boldsymbol{\xi}_\beta^\top \\
&\quad + \Delta_1\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\bar{\mathbf{M}}^-\Delta_1\boldsymbol{\xi}_\beta\bar{\boldsymbol{\xi}}_\beta^\top] \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} E[\bar{\boldsymbol{\xi}}_\alpha(\Delta_1\boldsymbol{\xi}_\alpha, \bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\beta)\Delta_1\boldsymbol{\xi}_\beta^\top + \bar{\boldsymbol{\xi}}_\alpha(\Delta_1\boldsymbol{\xi}_\alpha, \bar{\mathbf{M}}^-\Delta_1\boldsymbol{\xi}_\beta)\bar{\boldsymbol{\xi}}_\beta^\top + \Delta_1\boldsymbol{\xi}_\alpha(\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\beta)\Delta_1\boldsymbol{\xi}_\beta^\top \\
&\quad + \Delta_1\boldsymbol{\xi}_\alpha(\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^-\Delta_1\boldsymbol{\xi}_\beta)\bar{\boldsymbol{\xi}}_\beta^\top] \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} E[(\Delta_1\boldsymbol{\xi}_\alpha, \bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\beta)\bar{\boldsymbol{\xi}}_\alpha\Delta_1\boldsymbol{\xi}_\beta^\top + (\Delta_1\boldsymbol{\xi}_\alpha, \bar{\mathbf{M}}^-\Delta_1\boldsymbol{\xi}_\beta)\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top + (\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\beta)\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\beta^\top \\
&\quad + \Delta_1\boldsymbol{\xi}_\alpha(\bar{\mathbf{M}}^-\Delta_1\boldsymbol{\xi}_\beta, \bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\beta^\top] \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} E[\bar{\boldsymbol{\xi}}_\alpha((\bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\beta)^\top\Delta_1\boldsymbol{\xi}_\alpha)\Delta_1\boldsymbol{\xi}_\beta^\top + \mathrm{tr}[\bar{\mathbf{M}}^-\Delta_1\boldsymbol{\xi}_\beta\Delta_1\boldsymbol{\xi}_\alpha^\top]\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top + (\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\beta)\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\beta^\top \\
&\quad + \Delta_1\boldsymbol{\xi}_\alpha(\Delta_1\boldsymbol{\xi}_\beta^\top\bar{\mathbf{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\beta^\top] \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \Big(\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top\bar{\mathbf{M}}^- E[\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\beta^\top] + \mathrm{tr}[\bar{\mathbf{M}}^- E[\Delta_1\boldsymbol{\xi}_\beta\Delta_1\boldsymbol{\xi}_\alpha^\top]]\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top
\end{aligned}
$$

$$
\begin{aligned}
&+(\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^- \bar{\boldsymbol{\xi}}_\beta) E[\Delta_1 \boldsymbol{\xi}_\alpha \Delta_1 \boldsymbol{\xi}_\beta^\top] + E[\Delta_1 \boldsymbol{\xi}_\alpha \Delta_1 \boldsymbol{\xi}_\beta^\top] \bar{\mathbf{M}}^- \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\beta^\top \Big) \\
&= \frac{\sigma^2}{N^2} \sum_{\alpha,\beta=1}^{N} \Big(\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\beta^\top \bar{\mathbf{M}}^- \delta_{\alpha\beta} V_0[\boldsymbol{\xi}_\alpha] + \mathrm{tr}[\bar{\mathbf{M}}^- \delta_{\alpha\beta} V_0[\boldsymbol{\xi}_\alpha]] \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\beta^\top + (\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^- \bar{\boldsymbol{\xi}}_\beta) \delta_{\alpha\beta} V_0[\boldsymbol{\xi}_\alpha] \\
&\quad + \delta_{\alpha\beta} V_0[\boldsymbol{\xi}_\alpha] \bar{\mathbf{M}}^- \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\beta^\top \Big) \\
&= \frac{\sigma^2}{N^2} \sum_{\alpha=1}^{N} \Big(\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top \bar{\mathbf{M}}^- V_0[\boldsymbol{\xi}_\alpha] + \mathrm{tr}[\bar{\mathbf{M}}^- V_0[\boldsymbol{\xi}_\alpha]] \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + (\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^- \bar{\boldsymbol{\xi}}_\alpha) V_0[\boldsymbol{\xi}_\alpha] + V_0[\boldsymbol{\xi}_\alpha] \bar{\mathbf{M}}^- \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top \Big) \\
&= \frac{\sigma^2}{N^2} \sum_{\alpha=1}^{N} \Big(\mathrm{tr}[\bar{\mathbf{M}}^- V_0[\boldsymbol{\xi}_\alpha]] \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + (\bar{\boldsymbol{\xi}}_\alpha, \bar{\mathbf{M}}^- \bar{\boldsymbol{\xi}}_\alpha) V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[V_0[\boldsymbol{\xi}_\alpha] \bar{\mathbf{M}}^- \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top] \Big)
\end{aligned} \tag{B.1}
$$

References

Albano, R., 1974. Representation of digitized contours in terms of conics and straight-line segments. Comput. Graph. Image Process. 3, 23–33.
Al-Sharadqah, A., Chernov, N., 2009. Error analysis for circle fitting algorithms. Electron. J. Stat. 3, 886–911.
Amemiya, Y., Fuller, W.A., 1988. Estimation for the nonlinear functional relationship. Ann. Stat. 16, 147–160.
Bookstein, F.J., 1979. Fitting conic sections to scattered data. Comput. Graph. Image Process. 9, 56–71.
Cheng, C.-L., Kukush, A., 2006. Non-existence of the first moment of the adjusted least squares estimator in multivariate errors-in-variables model. Metrika 64, 41–46.
Chernov, N., Lesort, C., 2004. Statistical efficiency of curve fitting algorithms. Comput. Statist. Data Anal. 47, 713–728.
Chojnacki, W., Brooks, M.J., van den Hengel A, Gawley, D, 2000. On the fitting of surfaces to data with covariances. IEEE Trans. Pattern Anal. Mach. Intell. 22, 1294–1303.
Cooper, D.B., Yalabik, N., 1976. On the computational cost of approximating and recognizing noise-perturbed straight lines and quadratic arcs in the plane. IEEE Trans. Comput. 25, 1020–1032.
Fitzgibbon, A., Pilu, M., Fisher, R.B., 1999. Direct least square fitting of ellipses. IEEE Trans. Pattern Anal. Mach. Intell. 21, 476–480.
Gander, W., Golub, H., Strebel, R., 1995. Least-squares fitting of circles and ellipses. BIT 34, 558–578.
Gnanadesikan, R., 1977. Methods for Statistical Data Analysis of Multivariable Observations, 2nd ed., Wiley, Hoboken, NJ, U.S.A.
Hartley, R., Zisserman, A., 2004. Multiple View Geometry in Computer Vision, 2nd ed., Cambridge University Press, Cambridge, U.K.
Kanatani, K., 2006. Ellipse fitting with hyperaccuracy. IEICE Trans. Inf. Syst. E89-D, 2653–2660.
Kanatani, K., 1993. Geometric Computation for Machine Vision. Oxford University Press, Oxford, U.K.
Kanatani, K., 1996. Statistical Optimization for Geometric Computation: Theory and Practice. Elsevier, Amsterdam, the Netherlands, Reprinted, 2005, Dover, New York, U.S.A.
Kanatani, K., 2008. Statistical optimization for geometric fitting: theoretical accuracy analysis and high order error analysis. Int. J. Comput. Vis. 80, 167–188.
Kanatani, K., Sugaya, Y., 2007. Performance evaluation of iterative geometric fitting algorithms. Comput. Statist. Data Anal. 52, 1208–1222.
Kanatani, K., Sugaya, Y., 2008. Compact algorithm for strictly ML ellipse fitting. In: Proc. 19th Int. Conf. Pattern Recognition.
Kanatani, K., Sugaya, Y., 2010. Unified computation of strict maximum likelihood for geometric fitting. J. Math. Imaging Vision. 38, 1–13.
Kåsa, I., 1976. A curve fitting procedure and its error analysis. IEEE Trans. Instrum. Meas. 25, 8–14.
Kukush, A., Markovski, I. Van Huffel, S., 2004. Consistent estimation in an implicit quadratic measurement error model. Comput. Statist. Data Anal. 47, 123–147.
Leedan, Y., Meer, P., 2000. Heteroscedastic regression in computer vision: problems with bilinear constraint. Int. J. Comput. Vis. 37, 127–150.
Matei, B.C., Meer, P., 2006. Estimation of nonlinear errors-in-variables models for computer vision applications. IEEE Trans. Pattern Anal. Mach. Intell. 28,, 1537–1552.
Paton, K.A., 1970. Conic sections in chromosome analysis. Pattern Recognit. 2, 39–40.
Porrill, J. 1990. Fitting ellipses and predicting confidence envelopes using a bias corrected Kalman filter. Image Vis. Comput. 8, 37–41.
Pratt, V., 1987. Direct least-squares fitting of algebraic surfaces. Comput. Graph. 21, 145–152.
Rangarajan, P., Kanatani, K., 2009. Improved algebraic methods for circle fitting. Electron. J. Stat. 3, 1075–1082.
Rosin, P.L., 1993. A note on the least squares fitting of ellipses. Pattern Recognit. Lett. 14, 799–808.
Rosin, P.L., West, G.A.W., 1995. Nonparametric segmentation of curves into various representations. IEEE. Trans. Pattern Anal. Mach. Intell. 17, 1140–1153.
Taubin, G., 1991. Estimation of planar curves, surfaces, and non-planar space curves defined by implicit equations with applications to edge and range image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 13, 1115–1138.